GeoMx™ DSP Data Analysis Guidelines

GeoMx DSP Introduction

GeoMx Digital Spatial Profiling (DSP) is a new technology which enables spatial analysis of the immune contexture in tissue specimens on slides. Samples are stained with large panels of pre-mixed biological probes (each which incorporate a unique, UV-cleavable DNA barcode) and fluorescently-labeled morphology markers (used to visually or computationally elucidate the morphology of the tissue). Using the fluorescent morphology markers for guidance, defined regions of interest (ROIs) are illuminated with UV light and the cleaved barcodes from these ROIs are then counted on an nCounter platform. This allows for high-plex protein quantitation from spatially-resolved regions within the tissue. The resulting counts constitute an expression profile of key targets across an ROI and elucidate the biology specific to that region.

Biomarker Discovery

One common application of DSP, biomarker discovery, can potentially identify proteins whose expression levels in the tissue predict clinical response to immune-targeted therapeutics. This is particularly important to predict response to immunotherapy for patients with cancer, to optimize therapeutic regimens and to avoid adverse side effects. Biomarkers and targets are often differentially expressed in distinct regions of tumors (tumor center, invasive margin, stroma, lymphoid structures) and in biological segments within those regions (cancer cells, immune cells). Spatial analysis reveals and quantifies these biomarkers, but also allows for enrichment of segments within tumor microenvironments. Experimental designs supporting biomarker discovery applications typically involve analysis of various segments within the tumor from patients that fall into different clinical groups, such as treated/untreated, responder/non-responder, or relapse/non-relapse. With such designs, one can answer questions such as, *"Are there markers or combinations of markers that change in expression in specific biological segments of the tumor between patients that respond to immune checkpoint inhibitors (ICI) versus those that do not respond?"*. This document provides guidance on the analysis of DSP data in this experimental design context.

Experimental Design for Biomarker Discovery

In a biomarker discovery experiment, tumors are typically sampled and preserved in Formalin Fixed Paraffin Embedded (FFPE) blocks for sectioning on slides and studied with morphological stains, immunohistochemistry (IHC), or immunofluorescence. For DSP analysis, sectioned slides are stained with a panel of 20-50 oligo-tagged antibodies and two fluorescently-labeled antibodies (to distinguish the tumor by pan-Cytokeratin [PanCK] and the immune infiltrated regions [CD45]) and a fluorescent DNA dye. A fourth color channel is available for the user to add their own morphology or population marker. Contained within the panel of oligo labeled antibodies are three negative isotype controls (mouse IgG2a, Mouse IgG1 and Rabbit IgG1) and two antibodies suggested for positive controls of cellularity (Histone H3 [H3] and Ribosomal Protein S6 [S6]). Once stained, each slide is imaged, using the three markers to identify biologically-interesting areas in the tumor. It is then automatically segmented between tumor, based on PanCK-positive expression, and the tumor microenvironment (TME), the non-PanCK regions, as shown in Figure 1.

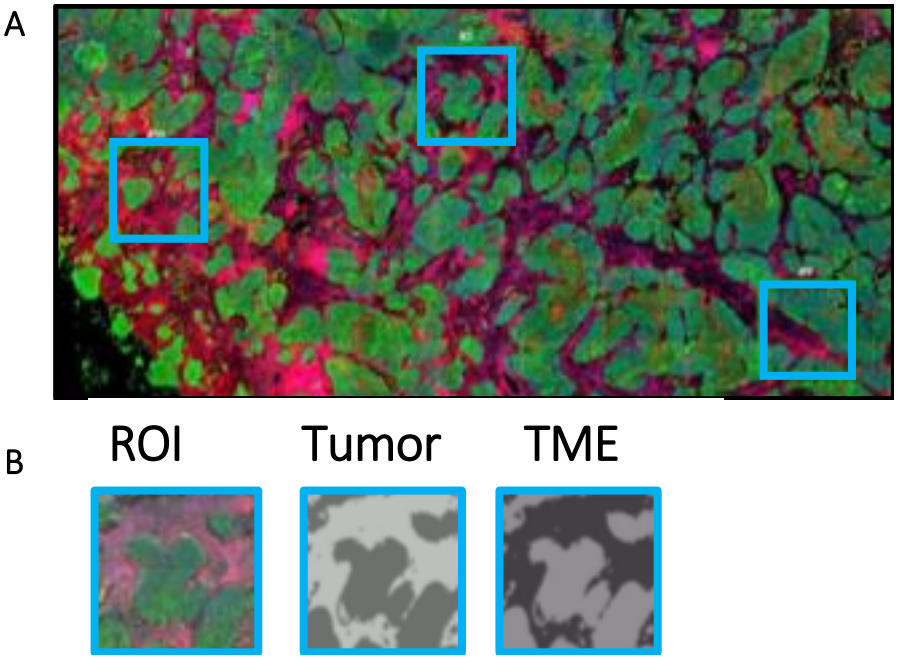

Figure 1: ROI selection strategy for Tumor/TME characterization. A) Square ROIs are established on a DSP scan based on fluorescent morphology markers, PanCK is shown in green, CD45 is shown in red. B) Within each ROI, segmentation is performed based on PanCK-positive (green=tumor) and PanCK-negative (not green=TME) visual morphology markers.

The result of this approach is the generation of an nCounter digital count profile for each DNA-barcoded antibody in two segments (tumor and TME) for every selected ROI within each sample and across multiple samples. Each digital count profile, segment, ROI and slide should be inspected for data quality throughout the normalization process. Once the data is properly QC'd and normalized, it can be queried for biology insight.

Data Analysis Workflow for Biomarker Discovery

As described in the introduction, the experimental example that is addressed in this analysis guideline is the discovery of proteins that are differentially expressed within the tumor (PanCK+) and TME (PanCK-) compartments of FFPE samples collected from responders and non-responders to an immune-targeted therapeutic. The hypothesis is that biomarkers specific to the tumor or TME compartment might predict response to therapy.

To address this hypothesis, we ask not how proteins are differentially expressed between the different segments within a tumor but how each segment differentially expresses markers between patient groups with two different outcomes. The steps below outline a general workflow for this type of study.

1. Normalize all data based on the ERCC controls. This is done automatically by the DSP Data Analysis Suite software during the QC step. Figure 4 in **Data Normalization** illustrates how to normalize your

own data. Normalization factors far outside a typical range (above 3.0) signify low quality data and indicate a profile that should be considered for removal.

2. Explore the various control probes and signal-to-background probes to identify and remove clear outliers. Examples of this include removing segments with high background or segments with low counts across all targets. The **Understanding Background** section has information on identifying outliers using these controls

3. In some cases, separate segment types may need to be normalized independently. Because tumor and TME are distinct, normalizing to the mean of both segment types could potentially result in extreme normalization factors and lack of biological clarity when considering differences. To consider segments separately, filter, then copy the data into another table.

4. When visualizing the clustering, determine if any of the segments are outliers to the rest of the segments of the same type. The dendrogram is very useful in determining how distinct these outliers are from the other segments. If the segment(s) are clearly outliers, they can be filtered out for (or from) further analysis. Importantly, analysis can be done with and, subsequently, without these outliers to understand their impact.

5. In order to determine that the appropriate normalization method has been used and outliers have been removed, each segment type (tumor and TME) should be separately analyzed between the two patient groups (responder and non-responder). The **Data Normalization** section contains information on each possible normalization step. You can apply one or more of these depending on the segments you selected and the type of tissue you are working with. This is an iterative process. A good normalization strategy will reveal both known and unknown biological differences.

6. For differential expression analysis, understanding the fold changes across the targets is the first step in analysis. A Mann-Whitney *U* test or Welch's T-test can then be used to look for significant differences between the two groups, depending on how skewed the distribution of probe expression is. The Mann-Whitney is appropriate if the data is not normally distributed and linear data should be $\log_2$ transformed before using the T-test and skewing should be inspected for each probe.

7. If multiple ROIs have been selected for analysis within a patient, more advanced tests, such as the Friedman test with repeated blocks or a mixed effect model, may be appropriate to account for the multiple sampling within a patient. However, to initially explore the data within the analysis software, an approximation can be made by calculating the geometric mean of the counts for each marker within each segment from a patient. For example, if there are 4 segments across a tumor, calculate the geometric mean of the counts observed in the 4 segments for each marker. This will result in one value per marker in the tumor, and one value per marker in the immune segments. After averaging, step 6 can be applied using the appropriate test for initial analysis.

8. Multiple test corrections should be used to provide greater stringency on the comparison, given the multiplicity of probes tested. We recommend starting with a false discovery adjustment, such as Benjamini-Yekutieli, if subsequent experiments will validate the findings or the results are exploratory. To identify the most robust changes between your groups of interest when

subsequent validation is not available, we recommend using a family-wise error adjustment, such as Bonferroni.

Understanding Background

All of NanoString's GeoMx DSP protein panels include isotype antibody controls (described in the following sections) which do not target specific proteins in human samples, thereby estimating the non-specific background in the assay. Some samples may have high background across the section or within an individual region. This background can be caused by a number of factors, such as sample processing, tissue artifacts (e.g., folds) or characteristics of a sample (e.g., necrotic tissue) leading to high isotype control counts.

Isotype control counts are useful in understanding the performance of the assay, but it is not advisable to use them in performing background subtraction due to the inherent nature of Poisson-distributed data and the assumed independent binding of each probe. In practice, background subtraction interferes with interpretation of low expression targets and has a negligible effect on mid and high expression targets. As seen in Figure 2, comparing Histone H3 counts and isotype controls can help identify outlier segments. Segments with high isotype control counts due to clear tissue artifacts should be removed from the analysis.

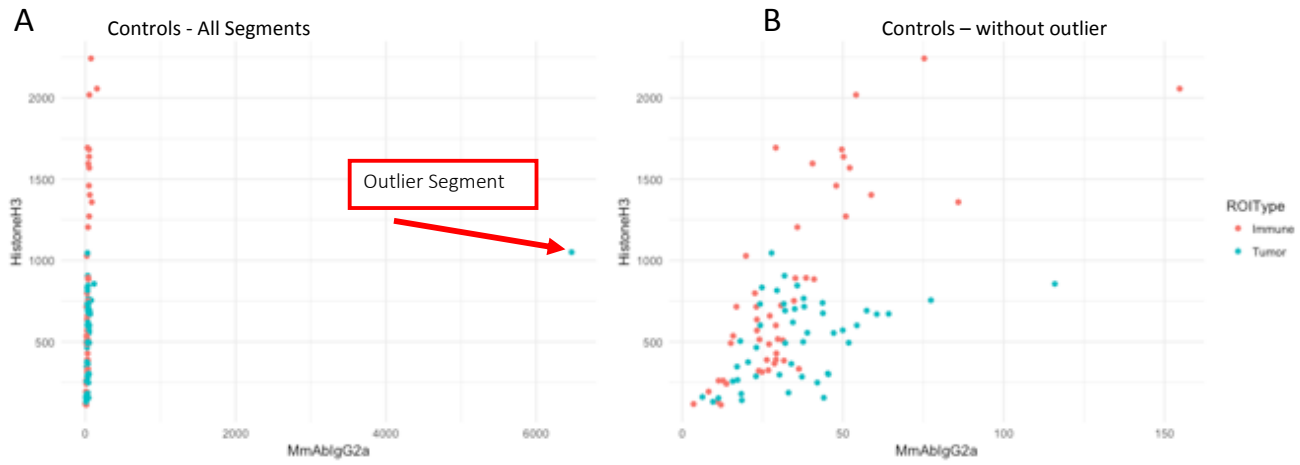

Figure 2: Background control QC metrics for a dataset. A) Linear comparison of Histone H3 counts and isotype control counts identifies a segment outlier. B) Removal of the outlier results in a clear picture of the dataset.

Below, in the Background Normalization and Transformation section, we will discuss how examining negative isotype control counts and their relationships to the other targets can offer more insight into assay quality. Note that it is possible that isotype control antibody binding can be higher than expected in some tissue types and should be interpreted within the context of the broader experimental results. Figure 3 compares signal-to-background ratios with negative control counts to illustrate the range of counts for

different antibody targets. In general, counts and the signal-to-background ratios should correlate well across similar types of segments (e.g., all tumor should be similar and all TME should be similar).

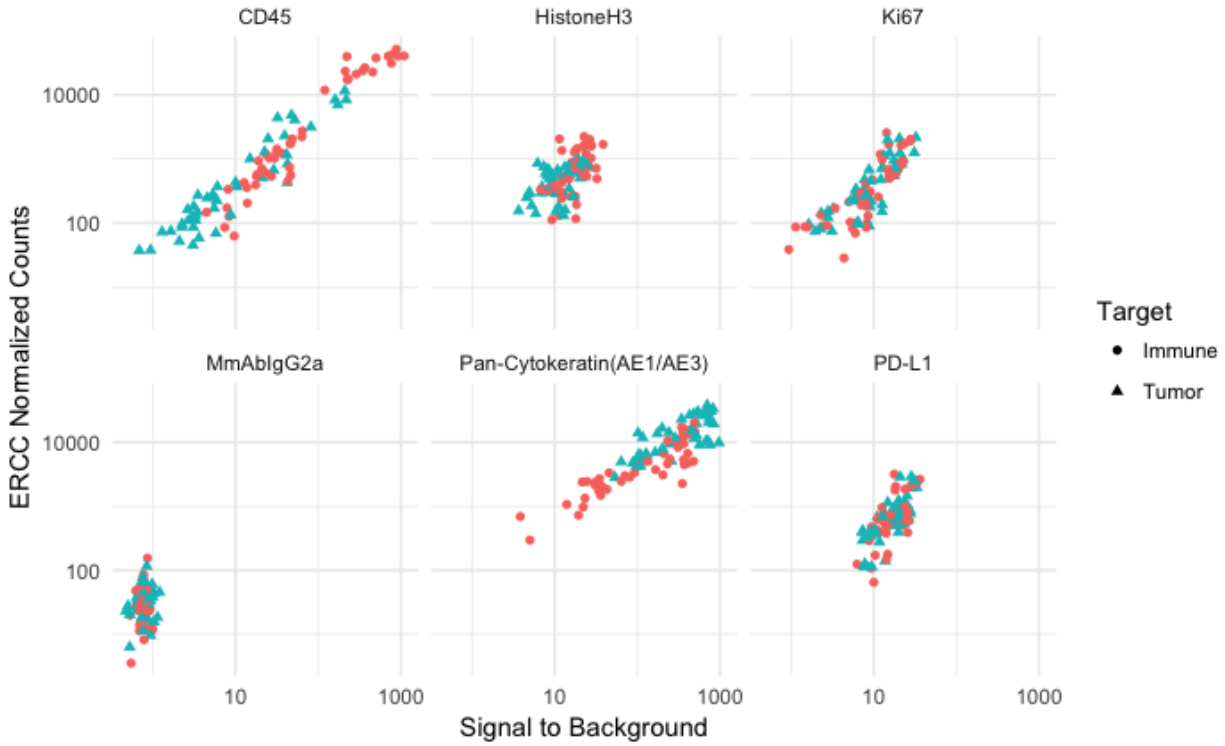

Figure 3: Signal-to-background for various targets.

Data Normalization

Normalization is a key step in accounting for technical variation and uncovering true biological changes. When analyzing similar segment types, normalization is particularly effective in accounting for differences in the area, cell density, staining stickiness, and oligo counting variation within each segment. It will not, however, remove the variation from widely different tissues (e.g. different tumor types) or segment types; these types of comparisons require careful controls and are best studied by comparing within sample ratios. For example, the difference between responders and non-responders in CD4 counts will be more informative than simply comparing the counts themselves between highly differential sample types.

Most normalization methods for count data are relatively straightforward and can be calculated in a spreadsheet.

1. Calculate the geometric mean for all of the normalization markers in each profile.
2. Calculate the average of the geometric means across all of the lanes.
3. Divide this average by the geometric mean in each lane to get a sample-specific normalization factor.
4. Multiply all of the target counts in each lane by the lane-specific normalization factor.

This process is illustrated below for the ERCC Controls but can be applied for any normalization marker. Be aware that different computational engines (R, Excel, python etc.) will give slightly different results, but the final counts should be very similar. Furthermore, normalization is good way to identify outliers. Profiles with normalization factors that deviate from the mean more than 4 standard deviation may be inappropriate to include in the analysis and should be interpreted cautiously if they are not removed from analysis.

ERCC Controls

Varied concentrations of External RNA Control Consortium (ERCC) synthetic oligos are in the GeoMx Hyb Code Packs to control for any technical variation that may occur in hybridization. It is by normalizing to these technical controls that we are able to account for the cumulative variation in raw counts across the samples, which arises from differences in pipetting errors, hybridization efficiencies and post-hybridization processing in the nCounter system. It is highly recommended that raw counts are normalized to the positive ERCC control probe counts to adjust for these inherent technical variabilities. This step is performed automatically by the DSP Data Analysis software. The normalization factors can be assessed to determine any technical issues with the nCounter processing and are included in QC.

A

	A	B	C	D
1	Original Counts			
2	**Column1**	**Sample 1**	**Sample 2**	**Sample 3**
3	ERCC control - POS A	33823	34387	30879
4	ERCC control - POS B	4107	4094	3717
5	ERCC control - POS C	1287	1233	1171
6	Sample Mean	5633.42	5578.30	5122.37
7				
8	Overall Average	5444.69		
9	Normalization Factor (Overall Average/Samples Mean)	1.0	1.0	1.1
10				
11	Normalized Counts			
12		**Sample 1**	**Sample 2**	**Sample 3**
13	ERCC control - POS A	32689.9	33563.4	32822.1
14	ERCC control - POS B	3969.4	3995.9	3950.9
15	ERCC control - POS C	1243.9	1203.5	1244.7

B

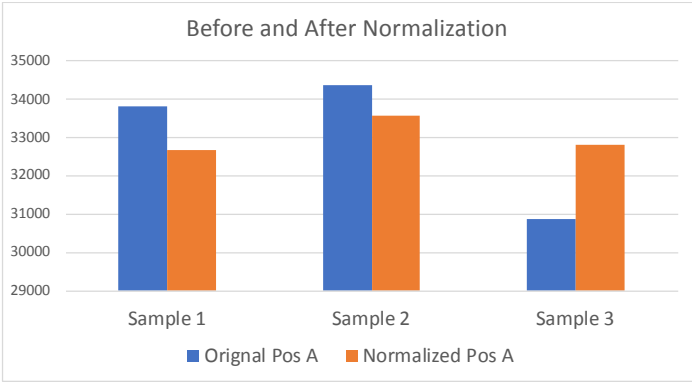

Figure 4: Normalizing to the ERCC controls. A) Example ERCC geomean normalization calculated in a spreadsheet. B) Comparison of Original Pos A to Normalized Pos A over multiple samples.

Housekeeper Normalization

Histone H3 and ribosomal protein S6 are included in the antibody mix as housekeeping markers. Normalization to housekeepers is recommended when comparing common segment types to account for differences due to the variation in the number of cells in a segment and/or the variation in sample processing methods (antigen retrieval, antibody cocktail, etc.) Housekeepers, however, can vary in different cell-types and may, therefore, not be appropriate in comparing data across different ROI or segment types.

Area Scaling

The DSP system automatically determines the surface area of each segment. Presumably, the tissue area will correspond to the number of cells, which, in turn, should correspond to the UV-cleavable probe counts. This assumption should be tested by comparing the area to other markers of global content such as nuclear count, Histone H3, etc. In this way, surface area provides an effective measure with which to normalize across segments of variable sizes.

Nuclear Counting (Not always available)

Using the DNA dye image, the DSP system automatically counts the nuclei as a surrogate marker for the number of cells within each segment. The cell count estimate from the DSP cell count algorithm is sensitive to the quality of the tissue, the biology (e.g., polynuclear cells), the staining, as well as the imaging, and should be evaluated independently for each specimen and area of interest.

Background Normalization and Transformation

By dividing each count in a profile by the geometric mean of the isotype negative controls, the signal strength of a particular target can be assessed. It is a useful method to assess the performance of the assay, as it provides a high-level view of positive signal for the specific antibodies in the panel relative to the non-specific binding that is inherent in the methodology. Reduced signal-to-background due to high background antibody counts can occur if the tissue region is particularly “sticky” or if the tissue section varies in thickness across the slide. As a general rule-of-thumb, signal-to-background ratios of less than 3 should be interpreted cautiously and should be subjected to additional downstream validation. Since isotype control counts scale with general stickiness, you can normalize to isotype controls to account for widely different sample quality. Alternatively, you can use the signal-to-background values in your analysis but note that the values may not compare easily across slides.